

ABSTRACT

A method for markup language document compression comprises defining a schema that specifies the structure of the markup-language conforming document and defining in the schema the types of elements and attributes that comprise the conforming document. The method further comprises assigning names in the schema for each of the elements and attributes of the document, defining relationships between the elements and between the attributes and the elements. The method further comprises assigning a token in the schema representing each element name and each attribute name of the document, and replacing each element name and each attribute name in the document with the assigned token.